

Defining the Limits of Grammatical Borrowing

Yaron Matras

University of Manchester

An examination of Romani and Domari, two Indo-Aryan languages spoken outside of India, shows that borrowing has an impact on a wide range of grammatical domains and categories. This includes even the domain of bound, inflectional morphology, often thought of as relatively immune to contact influence. Nonetheless, borrowing is not random, but tends to be structured in a hierarchical manner and so it is at least to some extent predictable. This suggests that, to the historical-comparative linguist, some components of grammar offer more reliable indicators than others about shared historical-genetic inheritance.

1. Introduction

This paper attempts to contextualise contact-induced change in the domain of grammar – ‘grammatical borrowing’ – by examining two case studies: The European dialects of Romani, and Domari, both territorially isolated New Indo-Aryan languages. These languages stretch both grammatical and lexical borrowing almost to their very limits, yet there are boundaries to the extent of borrowing even in these cases. In particular, it will be shown that contact influences in grammar do not occur at random, but tend to follow rather hierarchical patterns that are to some extent at least predictable. Given the general orientation of the volume toward aspects of historical reconstruction and the less contested observation that clause-level syntax is a weak indicator of genetic relatedness, my focus will be on morphology and grammatical (morpho-) lexicon.

2. The languages

Both Romani and Domari are descendents of Central Indo-Aryan languages dating back to the transition period from Old Indo-Aryan (OIA) to Middle Indo-Aryan (MIA). Both languages share their most ancient layer of innovations with Central Indo-Aryan: Syllabic *r* in OIA *śyn-* ‘to hear’ develops in both languages into a raised vowel (Romani *šun-*, Domari *sin-*). The cluster in OIA *akṣi* ‘eye’ is simplified to *k* (Romani *jakh*, Domari *iki*), and the cluster in OIA *asmnan*, *tusme* ‘we, you.PL’ loses its fricative segment (Romani *amen*, Domari *eme* ‘we’, Romani *tumen*, Domari *itme* ‘you.PL’). The initial semi-vowel in OIA *yuvatīḥ* ‘woman’ becomes an affricate (Romani *džuvēl*, Domari *džuwir*). This combination of isoglosses places the ancestral forms of both languages in close proximity to those of the present-day Central languages of India (cf. Masica 1991).

Both languages also preserve a number of archaic OIA features, such as the clusters *st* in Romani *vast*, Domari *xast* ‘hand’, and *dr* in Romani *drakh*, Domari *drak* ‘grape’, as well as intervocalic dentals such as Romani *gelo*, Domari *gara* ‘gone’ (OIA *gata*), all features that are not retained in the modern Central languages. Both languages also preserve, by and large, the consonantal present-tense person concord set, as in (1):

(1) Present-tense person concord (analogous formations and innovations are shaded)

	OIA	Romani	Domari
1SG	<i>-ami</i>	<i>-av(a)</i>	<i>-ami</i>
2SG	<i>-asi</i>	<i>-es(a)</i>	<i>-ēk</i>
3SG	<i>-ati</i>	<i>-el(a)</i>	<i>-ari</i>
1PL	<i>-amas</i>	<i>-as(a)</i>	<i>-ani</i>
2PL	<i>-ati</i>	<i>-en(a)</i>	<i>-asi</i>
3PL	<i>-anti</i>	<i>-en(a)</i>	<i>-andi</i>

Alongside these conservative traits, which indicate a rather early migration away from the Central regions, both languages participate in other changes that are typical of the transition period from MIA to New Indo-Aryan (NIA), such as reduction of gender and case marking in morphology, and the simplification of consonant clusters, in phonology (OIA *sarpa* ‘snake’, MIA *sappa*, Romani and Domari *sap*, cf. Hindi *sāp*).

Two rather diagnostic innovations are shared, to some extent at least, with languages of the extreme Northwest of India (e.g. Kashmiri), often referred to as the Indo-Iranian ‘frontier languages’ or ‘Dardic’ (cf. Grierson 1922). The first is the progression of grammaticalisation of postposed case particles into synthetic case markers, which in both languages is well-pronounced and has resulted in the emergence of not just one or two, but a whole series of five agglutinating suffixes (2):

(2) Romani and Domari Layer II case affixes

Case	Romani	Domari
Benefactive	<i>-ke/-ge</i>	<i>-ke</i>
Dative	<i>-te/-de</i>	<i>-ta</i>
Locative		<i>-ma</i>
Sociative	<i>-sa/-ca</i>	<i>-san</i>
Ablative	<i>-tar/-dar</i>	<i>-ki</i>
Genitive	<i>-ker-/-ger-</i>	

The second development is even more diagnostic of a close affinity to the ‘Dardic’ or North-western frontier languages in their early modern period. It concerns the emergence, post-ergativity, of a new past-tense conjugation, based on the attachment of oblique pronominal clitics to the past participle of the verb. The development is only partly complete, to different extents in each language (and with some variation within Romani). Typically, in

the third person the past tense of verbs of motion and change of state continues the old formation, in which the participle appears on its own, in an adjectival form, which agrees with the subject/undergoer: Romani *gel-o*, Domari *gar-a* ‘he went’, Romani *gel-i*, Domari *gar-i* ‘she went’. In transitive verbs, and in other persons also in all other verbs, the subject is represented by an historical oblique pronominal clitic, which has merged with the participle to form a new set of person concord markers (3):¹

(3) Romani and Domari past-tense person concord and its historical predecessors (for the verb ‘to do’) (analogies and innovations are shaded)

Person	MIA participle	MIA oblique pronoun	Romani	Domari
1SG	<i>ka(r)da</i>	<i>me</i>	<i>kerdjom</i>	<i>kardom</i>
2SG		<i>te</i>	<i>kerdjäl/-an</i>	<i>kardor</i>
3SG		<i>se</i>	<i>kerdjas</i>	<i>kardos</i>
1PL	<i>ka(r)de</i>	<i>ne</i>	<i>kerdjam</i>	<i>kerdēn</i>
2PL		<i>ve</i>	<i>kerdjan</i>	<i>kerdēn</i>
3PL			<i>kerde</i>	<i>karde</i>

The trail of archaisms, innovations and related isoglosses leads us to trace a very similar pathway in the development of the ancestral idioms of both Romani and Domari: Both seem to have broken away from an ancient Central Indo-Aryan cluster of dialects during the early transition period to MIA, but remained within reach of pan-Indo-Aryan innovations as late as the late MIA period and even the emergence of early New Indo-Aryan. While some of the conservatisms are idiosyncratic in the two languages, others are shared with languages of the Northwest, with which, in addition, some diagnostic late innovations are also shared.

In light of this very particular reconstructed history it is tempting to assume that Romani and Domari were both part of the very same lineage, and that they parted company only at a late stage in their development, possibly after leaving India. This was indeed the view taken by Sampson (1923) and other contemporaries. Turner (1926) however pointed out differences between the languages in the treatment of retroflex sounds (cf. Romani *ařo*, Domari *ata* ‘flour’, from *āṭa*; Romani *phuřo*, Domari *wida* ‘old man’, from *buḍḍha*), arguing for a rather old differentiation. Hancock (1995) has, in addition, pointed out that the inventory of Iranian loanwords in Romani and Domari is quite distinct, suggesting a split before Iranian influence and so before the exodus from Indian-speaking territories.

Outstanding differences also exist in grammatical vocabulary as well as in the position of pronominal affixes (which in Domari attach to nominal entities and prepositions, as well as to verbs). Other differences include the apparently rather early grammaticalisation in Romani of *-(j)ov-* (from OIA *bhuv-* ‘become’) as a mediopassive affix, while Domari retains OIA *-y-* (*bag-* ‘to break’, *bagy-* ‘to be broken’); the animacy constraint on the use of direct object marking in Romani (which is missing from Domari); the grammaticalisation of the verb *čh-* ‘to stay’ to a subjunctive marker in Domari intransitive verbs; the formation of the nominative plural in nouns (Romani *-a*, Domari *-e*); early reflexes of OIA *v-* and of OIA sibilants such as *ṣ* and *ś* (Romani *berš*, Domari *wars* ‘year’, OIA *variṣa*), and more (cf. Matras 1999).

The most likely historical scenario therefore seems to be an origin in a shared socio-ethnic rather than strictly linguistic community, with a shared migration history within India leading to participation in the same set of isoglosses, followed by a similar pattern of westwards migration out of India and subsequent retention of the original language under diverse foreign influences. In the socio-historical context of the Indian sub-continent, such a shared social identity that does not necessarily imply linguistic

unity can be fairly easily traced to the caste system with its numerous sub-divisions. This is confirmed by the employment in both communities of self-appellations deriving from the Indian caste-name *ḍom* (Romani *řom*, Domari *dom*), and moreover by the presence of a wholesale label for outsiders who are no caste-members, Romani *gadžo*, Domari *kaddža*, which has cognate expressions in the languages of present-day communities belonging to the *ḍom* and similar caste groups in India. The presence outside of India of other groups of Indian origin that carry caste-names as self-appellations, such as *Lom* in the Caucasus, *Jat* in Afghanistan or *Parya* in Tajikistan, confirms an overall phenomenon of emigration from India of specific caste groups and the maintenance of caste-like identity even after the breakaway from the actual caste-based social system of the Indian subcontinent.

In their subsequent history the two languages differ considerably. Earlier forms of Romani absorbed influences from Iranian and Armenian, as well as a very strong Byzantine Greek element, including many elements of vocabulary and even inflectional material. From the early fourteenth century onwards, Romani-speaking populations spread throughout Europe, leading to the dispersion and subsequent differentiation of the language into very distinct dialects. While Romani spread to virtually all regions of Europe, a high population density remained in central-eastern and South-eastern Europe. With well over 3.5 million speakers and possibly many more, Romani is now the largest minority language in the European Union. It is gradually occupying a space in the public domain, with numerous initiatives promoting literacy in the form of publications, websites, broadcasting and other media, as well as initiatives to introduce the language into the curriculum. Codification is generally region-based, with no uniform standard. Romani has been the subject of intense research dating back to the early eighteenth century, with an upsurge of interest in the modern period from 1990 onwards.ⁱⁱ

Domari, by contrast, is documented primarily in work on the variety spoken in Jerusalem (Macalister 1914, Matras 1999). Fragmented documentation of word lists and individual phrases allows us to reconstruct a picture of a spread of the language, probably in different varieties, from the Caucasus in the north and as far as Sudan in the south. It is reported that the language is still being spoken in communities in eastern Anatolia (Turkey), Lebanon and Syria, and Jordan. In most communities, it seems that use of the language is limited to the older generation. Jerusalem Domari is certainly highly endangered and is only spoken by a very small circle of elderly members of the community; the dominant family language is now Arabic. From Jerusalem Domari and other fragmented documentation of Palestinian Domari it appears that the language had been in close and prolonged contact with southern Kurdish, before immigration into Arabic-speaking territory. A layer of Turkish loanwords testify in all likelihood to the trade activities of the Dom as nomadic smiths and musicians under the Ottoman rule, but possibly also to close social links with nomadic Turkmen.

The following sections survey the contact behaviour of Romani and Domari in respect of selected functional categories. The comparison is biased by the existence of a large corpus of descriptions of Romani dialects, including the RMS (Romani Morpho-Syntax) Database with information on over 100 varieties of the language, which tags contact influences systematically (see Elšík & Matras 2006), while for Domari the material covers only the variety of a single location. Moreover, given the highly differentiated contact behaviour of Romani varieties in some domains, it is certainly possible that Jerusalem Domari may not be representative of Domari as a whole. Nonetheless, the enormous impact of contact both on Romani as a whole and on Jerusalem Domari, and the similar time-depth, socio-economic history, and ultimate genetic-typological origin of both languages justify a comparative discussion even in the absence of equal or even near-equal dialectological coverage.

3. Inflectional morphology

Inflectional morphology is well-known to be relatively resilient to borrowing, and therefore a rather stable indicator of genetic inheritance, though naturally it is subject to language-internal renewal and so it poses other challenges in the way of historical reconstruction. I therefore begin my survey of grammatical borrowing with this domain. The resistance of inflectional morphology to borrowing is best confirmed in the two languages in the area of case inflection: in neither of the two languages is there any direct attestation of borrowing of case markers. The closest to case marker borrowing is the apparent extension of what is a strongly lexicalised, inherited Romani marker of direction and location *-e* (e.g. *kher* ‘house’, vs. *kher-e* ‘home, at home’), in Zargari, a Balkan dialect of Romani that has migrated eastwards and is now spoken in Iranian Azerbaijan (Windfuhr 1970, Baghbidi 2003). Here, it is used much like a dative or directional case, possibly by analogy to the Azeri case marker *-e/-a*. The potential for the borrowing of case-markers is of course limited due to the paucity of transparent case markers in many of the contact languages. Those European languages that do have case inflection tend to possess highly flectional segments whose semantic transparency is rather low, a factor which inhibits borrowings (cf. Field 2002). Nonetheless, no borrowing is attested from languages like Finnish, Hungarian, Basque or Turkish either.

Borrowed nominal morphology is limited to nominative singular and plural endings. Domari incorporates Arabic nouns along with their plural ending, but adds its own plural formation to that: *zálame* ‘man’, plural *zlām-é* (from Arabic *zálame* ‘man’, plural *zlām*, with addition of the Domari plural ending *-e*). Romani shows productive use of borrowed nominal inflection endings. Early (Medieval) Romani borrowed Greek nouns along with their nominative inflection markers, thus *for-os* ‘town’ (from the Greek word for ‘market’), plural *for-i*, *kokal-o* ‘bone’, pl. *kokal-a*. These

Greek-derived inflectional endings continue to be productive in the European dialects of Romani and serve as the basis for the integration of further loanwords from contemporary contact languages: e.g. *prezident-os* ‘president’, *doktor-is* ‘doctor’, *šeft-o* ‘deal’ (German *Geschäft* ‘business’). Some individual Romani dialects continue to enrich the inventory of plural markers through borrowings from the contemporary contact languages. Vlax Romani adopts the Romanian-derived plural suffix *-uri/-ura*, which is employed both with earlier loans, e.g. *foruri* ‘towns’, and with new loans, as in *šefturi* ‘deals’ (but not combine with pre-European vocabulary). Epirus Romani, by contrast, adopts the Greek plural ending *-imata* which is applied to all inherited masculine nouns ending in a consonant: *vast* ‘hand’, PL *vastimata*; *kher* ‘house’, PL *kherimata*. Kaspichan Romani (northern Bulgaria) uses Greek-derived singular and plural endings with loans from Turkish: *džam-is* ‘mosque’ (Turkish *cami*), PL *džam-ides*.

Romani also shows borrowing in the domain of verb inflection. The Greek tense-aspect inflectional markers *-Vn-*, *-Vz-* etc. (present) and *-is-* (past) entered the language in all likelihood along with Greek-derived verbs, and were subsequently generalised to verbs from other contact languages. We find forms like *analadi-s-ker-djom* ‘I understood’, where the form *anladi* is the Turkish inflected past-tense 3SG form *anal-di* ‘understood’, the *-(i)s-* is the Greek aorist marker, *-ker-* is the inherited causative/transitive marker that integrates the loan verb, and *-djom* the inherited past-tense 1SG. These Greek markers have thus assumed a role within Romani morphology that is partly inflectional (indicating tense) and partly derivational (indicating particular verbs as loans).

Borrowed person concord markers on the verb appear marginally as well. Greek has contributed the 3SG present-tense marker *-i*. In many Romani dialects it is limited to loan verbs, e.g. Arli Romani of Kosovo *pomožin-i* ‘he/she helps’. In Slovene Romani it is generalised to all verbs, replacing the inherited Romani 3SG present-tense concord marker.

Slovene Romani also borrows the Slovene/Croatian person concord endings for the 2PL – *kerdž-ate* ‘you.PL did’, and the 1PL – *mothav-amo* ‘we say’. A cluster of Romani dialects in contact with Turkish, mainly in northern Bulgaria, borrow elements of their person concord set from Turkish, resulting in hybrid past-tense formations such as *kerd-amə̯s* ‘we did’ (Romani *kerd-am*, Turkish *yapt-ımız*) and *kerd-enə̯s* ‘you.PL did’ (Romani *kerd-en*, Turkish *yapt-ınız*). This process is triggered by analogies between the Romani 1st and 2nd person forms in *-m* and *-n*, respectively, and the corresponding Turkish forms in *-m* and *-n*, which is no doubt supported by the presence of similar nominal possession suffixes in Turkish. This analogy triggers a re-interpretation of the Turkish form *-Vz* as a generic plural marker, which is then attached to the inherited Romani forms (4):

(4) Borrowing of Turkish person concord markers into Romani

Person	Romani (inherited)		Turkish			Romani new form
	Present	Past	Present	Past	Nominal	Past
1SG		<i>-om/-em/-im</i>	<i>-Vm</i>	<i>-Vm</i>	<i>-Vm</i>	
2SG		<i>-an</i>	<i>-sVn</i>	<i>-Vn</i>	<i>-Vn</i>	
1PL	<i>-as</i>	<i>-am</i>	<i>-Vz</i>		<i>-VmVz</i>	<i>-am-ə̯s</i>
2PL	<i>-en</i>	<i>-en</i>	<i>-sVnVz</i>	<i>-VnVz</i>	<i>-VnVz</i>	<i>-en-ə̯s</i>

Even more widespread in the Balkans is the wholesale retention of Turkish verb inflection with lexical verbs borrowed from Turkish, resulting in the compartmentalisation of the verb inflectional domain into two separate sets of conjugations, inherited and borrowed. This can be illustrated by the contrast between the following two verbs in the Kalburdžu Romani dialect of Sindel, Northeastern Bulgaria (5):

(5) Inherited and Turkish-derived verb conjugations in the Kalburdžu Romani dialect of Sindel, Northeastern Bulgaria

	Inherited: <i>phurjo(v)</i> - ‘to grow old’		Turkish: <i>evlen-mek</i> ‘to marry’	
Person	Present	Past	Present	Past
1SG	<i>phurjovav</i>	<i>phurilem</i>	<i>evleniim</i>	<i>evlendim</i>
2SG	<i>phurjo</i>	<i>phurilan</i>	<i>evlenisin</i>	<i>evlendin</i>
3SG	<i>phurjol</i>	<i>phurila</i>	<i>evlenii</i>	<i>evlendi</i>
1PL	<i>phurjova</i>	<i>phurilam</i>	<i>evleniis</i>	<i>evlendik</i>
2PL	<i>phurjon</i>	<i>phurilen</i>	<i>evlenisinis</i>	<i>evlendinis</i>
3PL	<i>phurjon</i>	<i>phurile</i>	<i>evleniler</i>	<i>evlendiler</i>

The incipient use of original verb inflection with loan verbs can be observed in Russian Romani as well as in the Romani dialects of Greece. In the Greek Romani dialect of Epirus, for example, some Greek-derived verbs are adapted to Romani by means of loan verb adaptation markers – deriving, ironically, from Greek tense/aspect markers – but take inherited Romani inflection – *parakal-iz-ava* ‘I thank’ (Greek *parakal-o*), *aɣap-ez-ava* ‘I love’ (*aɣap-o*) – while others are inserted along with their Greek inflection:

(6)

na *bor-o* te *diavaz-o* soske *prep-i*
 NEG can-1SG COMP study-1SG because must-3SG

te *vojt-iz-av* me *daj-a*
 comp help-loan-1sg my-obl mother-obl

I cannot study because I have to help my mother’

The modal verbs *boro* ‘I can’ and *prepi* ‘must’ (impersonal) in (6) take Greek inflection, as does the lexical verb *diavazo* ‘I study’; but the Greek-derived lexical verb *vojt-* ‘to help’ is integrated into inherited Romani inflection.

The use of original, source-language inflection with borrowed modals is much more common than with borrowed lexical verbs, and is found in a number of Romani dialects. Gurbet Romani in Serbia, for example, replicates alongside the Serbian 3SG form *mora* ‘must’ also the Serbian 1SG *moram* ‘I must’ and 2SG *moraš* ‘you must’. In Domari, the borrowed Arabic modals *lāzim* ‘must’ and *mumkin* ‘it is possible’ are impersonal. But other modals and auxiliaries borrowed from Arabic take full Arabic person and tense inflection. They include the aspectual auxiliary *kān* ‘was’, which indicates past-habitual (*kān džari* ‘he used to go’, *kānat džari* ‘she used to go’, etc.), the transition-modals *šār* ‘to begin’ and *baqi* ‘to continue’, as well as the nominal form *bidd-ī* ‘I want’, *bidd-ak* ‘you.M want’, etc. It is noteworthy that the incorporation of Arabic inflection introduces a gender distinction into Domari in positions in which it is otherwise not indicated in the language (i.e. with present-tense verbs). The result is a compartmentalisation of the system of verbal morphology, separating most modal and auxiliary verbs, which are borrowed from Arabic and take Arabic inflection, from all lexical verbs (including those that are borrowed from Arabic), which take the inherited (Indic) inflection.

4. Deictic and anaphoric expressions

Deictic and anaphoric forms are rare among confirmed borrowings. Some Romani dialects of the Balkans have the place deixis *orde* ‘there’, possibly from Turkish or Azeri *orda/ordä*, as well as *inća* ‘here’, possibly from Persian *īndžā*, though both etymologies require further investigation. Some varieties of Sinti Romani borrow German-derived *doti* ‘there’. Rumungro (Selice)

Romani borrows the Hungarian deictic prefixes *am-* and *ugyan-*, which are combined with Romani deictic stems (Elšík 2007). Domari makes occasional use of the Arabic resumptive pronoun *īyyā-*, complete with its Arabic agreement inflection, in relative clauses. Domari also employs Arabic the reflexive expression *ḥāl-* ‘-self’ and the Arabic reciprocal *bašāq*.

A number of Romani dialects show the development of hybrid pronominal forms (7):

- (7) Borrowed pronominal forms in Romani and their origin:

	3SG	3PL	Source	3SG	3PL
Early Romani	<i>ov</i>	<i>on</i>			
Hungarian Romani	<i>ov</i>	<i>on-k</i>	Hungarian	<i>ő</i>	<i>ő-k</i>
Slovene Romani	<i>ov</i>	<i>on-i</i>	Slovene	<i>on</i>	<i>on-i</i>
Thracian Romani	<i>ov</i>	<i>on-nar</i>	Turkish	<i>o[n-]</i>	<i>on-lar</i>
Molise Romani	<i>jov</i>	<i>lor</i>	Italian	<i>il</i>	<i>loro</i>

Early Romani (see Matras 2002, Elšík & Matras 2006) is assumed to have had the pronominal forms *ov* ‘he’, *on* ‘they’, which are continued in most dialects of South-eastern Europe, while in other areas prothetic consonants *j-* and *v-* are added. The remarkable development attested in the top three dialects – Hungarian, Slovene, and Thracian Romani – is the copying of plural affixes from the contact languages into the inherited plural pronominal form. This is prompted by the accidental similarity of forms between the Romani pronouns and those of all three contact languages, as well as by the presence of an agglutinating formation

of the plural pronoun in the contact language, which is identical to the general, nominal plural suffix used in the language. It is thus the plural affix, not the actual pronominal form of the respective contact language that is borrowed into Romani. The exception is the borrowing of a word-form in Molise Romani, which is the only genuine case of the borrowing of a pronominal word-form. Molise Romani also borrows the reciprocal and 1PL reflexive form *ci* from Italian.

5. Derivational morphology

Romani dialects borrow a series of agentive and diminutive affixes from various contact languages. Common in particular are diminutive/diminutive feminine *-ic--ica/-icka-* and agentives *-ari* and Turkish-derived *-dži*, which are shared by several languages in the Balkans. Borrowing of other nominal derivation is less common. Borrowing of verb-deriving morphology is quite rare. A rather isolated example is Lovari Romani *-áz-* from Hungarian, found in a limited number of inherited words, such as *bučáz(in)-* ‘to work’, from *buči* ‘work’.

Romani dialects in contact with Slavic languages primarily in Poland, Russia, and Slovakia, and to some extent also Romani dialects in contact with Latvian, Lithuanian, and Greek, borrow *Aktionsart*-derivational prefixes (often referred to as ‘aspect’): In Russian Romani we find, based on *dava* ‘I give’, the verb *dodava* ‘I add’, *obdava* ‘I embrace’, *otdava* ‘I confiscate’, *piridava* ‘I hand over’, *podava* ‘I obtain’, *rozdava* ‘I hand out’, *vydava* ‘I give away’. In Latvian Romani we find, based on *dža-* ‘to go’, *iedža-* ‘to go in’ and *piedža-* ‘to approach’, and in some Greek Romani dialects we find *dikh-* ‘to see’, and *ksanadikh-* ‘to see again’.

Markers of adjective comparison in Romani often derive from pre-posed, unbound or semi-bound markers of the various contact languages, such as Slavic *po*, Romanian *mai*, and Turkish *daha* for the comparative, Slavic *naj*, Hungarian *leg-*, Turkish *en* for the superlative. A bound comparative/superlative marker *-eder*

had been borrowed into Proto-Romani from Iranian, and continues to be used in many dialects. Sinti Romani for example uses it for the comparative – thus *sik* ‘fast’, *sikedər* ‘faster’ – and adopts the German superlative, including both its synthetic ending and an accompanying preposition: *am sikestə* ‘fastest’, (dialectal) German *am schnell-ste(n)*.

There are no parallels to the borrowing of productive derivational morphology in Domari, possibly due to the highly inflectional character of Arabic, which makes it difficult to isolate individual morphemes as markers of a derivational procedure. The tendency is instead to borrow entire word-forms. This tendency has an interesting impact on the domain of adjective comparison. Arabic employs a morpho-phonological template *áCCaC* to derive comparative/superlative forms from consonantal roots: *kbīr* ‘big’, *ákbār* ‘bigger’; *zǧīr* ‘small’, *ázǧār* ‘smaller’. This template cannot easily be isolated or integrated into the agglutinative-inflectional morphological structure of Domari, nor is it simple or even possible to break down Domari adjectives such as *tilla* ‘big’ or *kištota* ‘small’ into tri-consonantal roots for insertion into the Arabic-based derivation template. The solution adopted by Domari speakers is to borrow the full Arabic word-form for all comparative/superlative forms, resulting in complete borrowing-based suppletion of the inventory of adjectives: *tilla* ‘big’, *ákbār* ‘bigger’; *kištota* ‘small’, *ázǧār* ‘smaller’.

6. Other morpho-lexicon

This section surveys a number of categories that are typically expressed in both languages by unbound, often uninflected function words. Romani borrows its numerals *efia*, *oxto*, *enja* (‘seven-nine’) as well as higher numerals above twenty (with the exception of ‘one hundred’) from Greek. Many dialects tend to replace these higher numerals through loans from the contemporary contact languages. Domari speakers in Jerusalem are unable to recall non-Arabic numerals above ‘five’, with the

exception of ‘ten’ and ‘hundred’, although the full set of numerals is documented for speakers in the same community in the early 1900s (cf. Macalister 1914).

Some varieties of Sinti Romani have borrowed the dialectal German negator *nit*, which follows the finite verb, as it does in German: *me džinau nit* ‘I don’t know’. Other Sinti varieties use the German-derived emphatic particle *gar*, roughly ‘indeed’, as a negation particle: *me džinau gar* ‘I don’t know’. This functionalisation is internal to Sinti, and appears to have derived via the German expression *gar nicht* ‘not at all’. Domari borrows the Arabic negator *mišš* in non-lexical predications – *pandži mišš mišt-ēk* ‘he *not* ill-COP.M.SG’ = ‘he is not ill’. The Arabic negator *mā* ... -š accompanies all Arabic-derived inflected verbs (modals and auxiliaries) in Domari: *pandži mā kānš mišt-ēya* ‘he was not ill’.

Romani dialects tend to borrow prepositions that express more peripheral and more complex local relations. These include relations that have complex reference points, such as ‘between’, ‘along’, ‘through’, and ‘around’, those that involve separation from a source, such as ‘from’, ‘toward’, ‘against’ and ‘since’, and especially those that convey contrast with a presupposed set, such as privative ‘without’ and ‘except’, and replacive ‘instead of’. Typical borrowings into various dialects include *pretiv/protiv* (Slavic) ‘against’, *is* (Slavic) and *fon* (German) ‘from’, *za* (Slavic) and *bis* (German) ‘until’, *de* (Romanian) ‘since’, *bez/brzo* (Slavic), *xoris* (Greek), *utan* (Swedish), and *oni* (German) ‘without’, *vmesto/namesto* (Slavic) ‘instead’, *osven, skluchenje, kromje* (Slavic), *in loc dθ* (Romanian) ‘instead’, and *ektos* (Greek) ‘except for’. The preposition ‘with’ is also borrowed in some dialects, e.g. Sinti Romani *mit* from German, Greek Romani *me* from Greek. Domari, by contrast, borrows most of its prepositions from Arabic. Core prepositions borrowed from Arabic are *maš* ‘with’, *min* ‘from’, *bašd* ‘after’, *šan* ‘about’, *šand* ‘at’, and *žamb* ‘next to’, in

addition to *badāl* ‘instead of’, *bala* ‘without’, *bēn* ‘between’. Some inherited (Indo-Aryan) adverbial expressions of local relations, such as ‘above’, ‘outside’, and ‘inside’, are nevertheless retained.

As discussed above, Romani dialects borrow modality word-forms such as Slavic and Romanian *musaj/musi-/muši-* and *triba/treba/trobu-* ‘must’, Slavic *mora* ‘must’ and *može* and *mog-* ‘can’, Turkish-derived *lazimi* and *medžburi* ‘must’, Swedish-derived *moste* ‘must’, German-derived *braux-* ‘need’, and Greek-derived *prep-* ‘must’ and *bor-* ‘can’ (see Matras 2002, Elšík & Matras 2006). Domari, too, borrows most of its modals and auxiliaries from Arabic – *lāzim* ‘must’ and *mumkin* ‘it is possible’, *kān* which indicates past-habitual, *šār* ‘to begin’, *baqi* ‘to continue’, *bidd-* ‘to want’ – an exception being the verb *sak-* ‘to be able to’.

All Romani dialects borrow indefinite markers, and many indefinite word-forms are also borrowed. Borrowed markers include Slavic (*v*)*sako*, *ni-*, *bilo-*, Romanian-derived *vare-*, Turkish-derived *hič*, *her*, *bazi*, Hungarian *vala-*, and more. Borrowed indefinite word-forms are numerous and include Slavic *ništo* ‘something, nothing, anything’, Polish *zawsze* ‘always’, Romanian-derived *mereu* ‘always’, Hungarian-derived *šoha* ‘ever, never’ and *mindig* ‘always’, Greek-derived *čipota* ‘nothing’ and *kathenas* ‘somebody’, and many more. Domari borrows the Arabic indefinite markers *kull* ‘every’, *ayy* ‘any’, and *wala* ‘no, none’, and combines them with the inherited Domari expression for person (*ekak*) and otherwise with Arabic expressions for thing, time, place, and so on. The word for ‘always’ is Arabic *dā?iman*.

Romani interrogatives are generally retained from the Indic stock of forms, though occasional borrowings are found in particular among the interrogatives of quantity (‘how much?’). Domari borrows both from Arabic – *qadēš?* ‘how much /many?’, and *waqtēš?* ‘when?’ – along with the determiner-interrogative *ayy?* ‘which?’. Subordinating particles in Romani are generally grammaticalised interrogatives, but word-forms for ‘because’,

‘although’, and ‘if’ are frequently borrowed, as are, occasionally, relative pronouns and the factual complementiser, for which we often find Greek *oti*, Bulgarian *či*, Romanian-derived *ke*, Italian *ke*, and Hungarian-derived *hodž/hod/hoi*. Romani dialects always borrow ‘but’ from the contemporary or recent contact language (e.g. Slavic *no*, *po* and *ali/ale*, Hungarian *de*, Turkish *ama*, Greek *ala*, German *aber*). Many Romani dialects also borrow ‘or’ and ‘and’. These latter two are borrowed frequently, but are sometimes retained from an older contact language. Ajia Varvara Romani in Athens for instance has *ja* ‘oder’ from its recent contact language Turkish, but *ala* ‘but’ from its current contact language Greek; Finnish Romani has *elle* ‘or’ from Swedish, but *mut* ‘but’ from Finnish; Manush Romani in France has German *un* ‘and’ and *otar* ‘or’, but French-derived *me* ‘but’, and so on (cf. again Matras 1998 & 2002, as well as Elšík & Matras 2006). Domari borrows all its connectivity devices from Arabic, including all coordinating and subordinating conjunctions, the relative particle *illi* and the factual complementiser *inn-*. Both languages always borrow discourse particles and interjections from the immediate contact language.

Finally, both languages borrow a considerable amount if not the full inventory of word-forms in the domains of phasal adverbs and focus particles. In Romani these are always loans from European languages, but they are relatively stable compared to other borrowed function words such as connectors or discourse markers. Thus some varieties of German Romani (Sinti) preserve Greek-derived *komi* ‘still’, but have German *schon* ‘already’ and *bloß* ‘only’, while Lovari Romani in Poland, Hungary, and Russia preserves Romanian-derived *inke* ‘still’, *aba* ‘already’, and *feri* ‘only’, and Burgenland Romani in Austria retains Hungarian-derived *meg* ‘still’, *imar* ‘already’, and *čak* ‘only’. Domari borrows all its relevant particles – *bass* ‘only’, *kamān* ‘too’, *ḥatta* ‘even’, and so on – from Arabic.

7. Discussion

The present contribution was intended to address the question of which grammatical categories, especially in morpho-syntax and morpho-lexicon are more prone to contact-induced change, and which are more reliable indicators of genetic-historical inheritance - and so, in comparative perspective, of language-genetic relatedness. The comparative or historical linguist approaching Romani and Domari will find an entire set of categories thoroughly unreliable as indicators of inheritance, and at the same time diagnostic of recent and fairly recent contacts leading to the cross-linguistic diffusion of word-forms. This includes discourse markers, connectors, phasal adverbs, focus particles, indefinite expressions and indefinite markers, expressions of modality (modal verbs and modal particles), and the marking of comparison in adjectives. Expressions of aspect, nominal derivation, plural formation, prepositions, lower numerals, negation and interrogatives constitute a mixed set that is penetrated by borrowings to some extent. The reasons for the susceptibility of these categories to borrowing is as functional as it is, quite possibly, facilitated by their relative structural autonomy (see discussion in Matras 2007 & 2008).

By contrast, case and possessive inflection, demonstratives and personal pronouns, and verb inflection are based, by and large, on inherited material. The categories are not, however, impenetrable to borrowings. Nonetheless, borrowing in these categories is not just quantitatively restricted, but it is also limited to particular patterns. In case inflection, the only evidence for contact-induced change alluded to above was the semantic extension of an inherited inflectional marker due to its resemblance to a functionally related marker in the contact language (Zargari Romani *-e*, in contact with Azeri). It is crucial, at this point, to distinguish two types of contact-induced change. The first, to which most of the present discussion has been devoted, is the direct replication of linguistic Matter or

phonological shapes of morphs along with their meanings or aspects thereof. The second is the replication of Pattern – the linear arrangement of morphs, or the meaning associated with them (cf. Matras & Sakel 2007, Matras 2008). Since our focus is on identifying inherited linguistic material, patterns have been largely excluded from this discussion.

In terms of Matter-replication, we therefore find nothing in the domain of case inflection. In pronouns, we saw the borrowing of plural markers in several instances, and the borrowing of word-forms in the case of Molisean Romani 3PL and reciprocal 1PL, and the Domari reciprocal and reflexive. Interestingly, these are all parts of the pronominal paradigm that are generally volatile. Romani dialects frequently renew the form of the 3PL (cf. Elšík & Matras 2006), and reciprocals and reflexive are frequently composed of secondary expressions, and so constitute recent grammaticalisations. A ‘core’ consisting of the deictic (first and second person) pronouns remains, in this survey, immune to contact. Demonstrative stems are generally inherited (though they may be augmented by borrowed markers), but here too we find considerable internal renewal and a great diversity in the formation and re-formation of stems (cf. Matras 2002).

In verb inflection, finally, with the only exception of three borrowed person affixes in Slovene Romani (one of them derived from Greek), the distribution of borrowed markers is constrained by an internal compartmentalisation of the system. Thus Greek-derived tense markers may accompany borrowed word-stems in Romani, but they do not serve as independent markers of tense and do not compete with or replace the inherited tense markers. The Greek-derived 3SG marker *-i* is found only in (some) borrowed verbs (except in Slovene Romani, where it is generalised). Use of verb inflection from the contact languages is otherwise characterised by a wholesale import of the entire paradigm, rather than of individual markers, accompanying either just modal verbs (in Domari and some Romani dialects, such as Gurbet), or modal verbs and a selection of borrowed lexical verbs (as in Epirus

Romani), or all borrowed verbs from the particular contact language (as in some Romani dialects in contact with Turkish). In all cases, the imported verb paradigm exists in complementary distribution alongside the inherited inflection paradigm.

The one additional case of a change to individual markers (exemplified above, for Romani from Sindel) can be explained, much like the case of 3PL pronouns in some Romani dialects, as a replication of a plural morpheme rather than the substitution of an entire person marker. Adding both cases together – the replication of plural formation in pronouns and in the person concord markers on the verb – we might generalise that plural inflection is more prone to borrowing. Indeed, the employment of borrowed plural markers on nouns provides further confirmation.

Even within the categories that show relative resilience to borrowing, we therefore find some borrowing, albeit tightly governed by a number of constraints: plural inflection markers are more likely to be borrowed independently of other markers in the paradigm, and full person concord paradigms may accompany borrowed lexical verbs and especially modals but are not likely to diffuse into inherited verbs. Thus while no domain of grammar, save perhaps case inflection, appears to enjoy general immunity toward borrowing of linguistic matter, inflection does show considerable stability in the two languages examined, and deictic and anaphoric paradigms are volatile but their renewal is driven by internal pressures and draws on internal (inherited) resources. The historical linguist aiming to reconstruct the history of Romani or Domari – two languages that have been in intense contact, as low-prestige, minority languages spoken exclusively by bilinguals – would encounter on the whole coherence between the source of core vocabulary and that of inflectional morphology, with disruptions being by and large recognisable as wholesale imports of paradigms along with borrowed lexicon.

Nonetheless, we must bear in mind that this finding draws on two important factors. The first is the relatively shallow time depth of contact developments. The radical compartmentalisation

in the verb system of some Romani, not to mention the complete replacement of the inventory of expressions in the more 'borrowable' categories, represents a time-depth of merely 500-600 years. The Greek impact on Romani is much older in terms of its absolute age, but considering that it receded with the breakup of the Byzantine Empire, we might again reconstruct a period of contact of somewhere between 400-500 years. Similarly, Domari has been in contact with Arabic possibly since the Saladin conquests in the twelfth century at the very earliest and in all likelihood for a much shorter period. Now, there is no reason to believe that contact-induced change should proceed at a steady pace, and so we should not assume that, had contact between Romani and Greek continued, then substantial additional components of the language might have changed. Indeed, we may gain some insights by examining those Romani dialects that are still spoken in Greece today. Nonetheless, both our sample languages represent periods of contact that are not immensely long; we cannot rule out the possibility that developments extending for millennia might have a stronger impact on a language.

The second factor is the nature of contact and its relevance to language maintenance. Both Romani and Domari are 'languages in contact', rather than 'contact languages'. This means that the contact influences that they show are a result of prolonged and gradual change. Even a comparison of Domari sources from the early twentieth century (Macalister 1914) with material collected recently in the same community (Matras 1999) shows some degree of gradual change among generations (in this case, the loss of inherited higher numerals during this time interval). The outcome can be very different in cases of an abrupt turnover of the balance of functions between the languages of a bilingual community, and the partial abandonment of a language within one or two generations. This is the case with Angloromani, the variety of British Romani that ceased to exist as an everyday family language toward the end of the nineteenth century, and was replaced instead

by a register that serves the function of an emotive discourse mode (cf. Matras et al. 2008). This functional turnover carries with it the abrupt loss of most grammatical inflection paradigms, especially finite verb inflection, and the reduction of the old language to, by and large, just a lexical inventory. Similar processes are attested from contemporary, ongoing cases of language shift, sometimes involving the partial retention of grammatical material (cf. O’Shannessy 2005, McConwell & Meakins 2005). It is now well-established in the contact linguistic literature that such cases of Mixed Languages (Bakker 1997, 2000) are not reliable indicators of genetic inheritance (cf. Thomason 1997). In the present contribution, I hope to have been able to illustrate that in languages that are not contact languages of the latter type, but have nevertheless been shaped by the huge impact of contact, a core of coherent linguistic-structural material representing an unambiguous source of genetic inheritance can still be identified.

Endnotes

ⁱ In Domari, this set also attaches to nouns to express possessive relations: *kury-om* ‘my house’, *kury-or* ‘your house’, *kury-os* ‘his/her house’, etc.

ⁱⁱ See Matras (2002) for an introduction and overview; for general information, bibliography, and on online database comparing Romani dialects in transcription and sound see the Manchester Romani Project website on <http://romani.humanities.manchester.ac.uk/>

References

- Baghbidi, H. R.
2003 The Zargari language. An endangered European Romani in Iran. *Romani Studies* 5/13: 123-148.
- Bakker, P.
1996 Language intertwining and convergence: typological aspects of genesis of mixed languages. *Sprachtypologie und Universalienforschung* 49: 9-20.
2000 Rapid language change: Creolization, intertwining, convergence. In: Renfrew, C., McMahon, A. & Trask, R. L. (eds), 585-620.
- Elšík, V. & Matras, Y.
2006 *Markedness and language change: The Romani sample*. Berlin: Mouton de Gruyter.
- Elšík, V.
2007 Grammatical borrowing in Hungarian Rumungro. In: Matras, Y. & Sakel, J. (eds), 261-282.
- Field, F. W.
2002 *Linguistic borrowing in bilingual contexts*. Amsterdam: John Benjamins.
- Grierson, G. A.
1922 *Linguistic survey of India*. Calcutta: Superintendent Government Printers.
- Hancock, I. F.
1995 On the migration and affiliation of the Dōmba: Iranian words in Rom, Lom, and Dom Gypsy. In: Matras, Yaron. (ed.), 25-51.
- Macalister, R. A. S.
1914 The language of the Nawar of Zutt, the nomad smiths of Palestine. *Gypsy Lore Society Monographs* 3. London: Edinburgh University Press.
- Masica, C. P.
1991 *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matras, Y. & Sakel, J.
2007 Investigating the mechanisms of pattern-replication In Language convergence. *Studies in Language* 31: 829-865.

- Matras, Y. & Sakel, J. (eds)
2007 *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton de Gruyter.
- Matras, Y.
1998 Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36: 281-331.
1999 The state of present-day Domari in Jerusalem. *Mediterranean Language Review* 11: 1-58.
2002 *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
2007 The borrowability of grammatical categories. In: Matras, Y. & Sakel, J. (eds), 31-74.
2008 *Language contact*. Cambridge: Cambridge University Press.
- Matras, Y. ed.
1995 *Romani in contact: the history and sociology of a language*. Amsterdam: Benjamins.
- Matras, Y., Gardner, H., Jones, C. & Schulman, V.
2008 Angloromani: A different kind of language? *Anthropological Linguistics*.
- McConvell, P. & Meakins, F.
2005 Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25: 9-30.
- O'Shannessy, C.
2005 Light Walpiri: A new language. *Australian Journal of Linguistics* 25: 31-57.
- Renfrew, C., McMahon, A. & Trask, R. L. (eds)
2000 *Time depth in historical linguistics*. Vol. 2. Cambridge: McDonald Institute for Archeological Research.
- Sampson, J.
1923 On the origin and early migrations of the Gypsies. *Journal of the Gypsy Lore Society*, third series, 2: 156-169.
- Thomason, S. G.
1997 A typology of contact languages. In: Spears, A. A. & Winford, D. (eds) *The structure and status of pidgins and creoles*. Amsterdam: John Benjamins. 71-88.
- Turner, R. L.
1926 The position of Romani in Indo-Aryan. *Journal of the Gypsy Lore Society*, third series, 5: 145-189.
- Windfuhr, G.
1970 European Gypsy in Iran: A first report. *Anthropological Linguistics* 12: 271-292.